RESEARCH ARTICLE

BORNEO SCIENCE
THE JOURNAL OF SCIENCE & TECHNOLOGY

Journal homepage: https://jurcon.ums.edu.my/ojums/index.php/borneo-science

IN SILICO ANALYSIS AND STRUCTURE MODELLING OF GAHP2, A CONSERVED HYPOTHETICAL PROTEINS RELATED TO THERMAL STRESS RESPONSE IN *GLACIOZYMA ANTARCTICA* PI12

Makdi Masnoddin[1], Clemente Michael Wong Vui Ling[2] and Nur Athirah Yusof [2,*]

[1] Preparatory Center for Science and Technology, Universiti Malaysia Sabah, Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia.
[*2] Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia.

Correspondence:
nrathirah.yusof@ums.edu.my

Received: 10 April 2025
Revised: 20 May 2025
Accepted: 22 May 2025
Published online: 27 May 2025

DOI :
10.51200/bsj.v46i1.6314

Keywords:
Antarctic yeast; Thermal stress response; Homology modelling; Structural and functional annotation.

ABSTRACT. *The genomic data of the native Antarctic yeast, Glaciozyma antarctica PI12, has garnered attention due to its distinctive thermal adaptation. Nonetheless, a significant percentage of the proteins associated with thermal stress adaptation were identified as conserved hypothetical proteins (HPs), suggesting that these proteins remain experimentally uncharacterized. Consequently, this study aims to determine the structural characteristics of GaHP2, an uncharacterized conserved hypothetical protein believed to play a significant role in the thermal stress response. The gene was subjected to an extensive analysis utilizing computational tools to explore its function, physicochemical properties, and three-dimensional structure. Functional annotation was executed using NCBI BLAST and InterProScan; physicochemical properties were evaluated with ExPASy's ProtParam; homology modeling was performed using Phyre2 and AlphaFold2, while structure validation, refinement, and superimposition were implemented with ModRefiner and UCSF Chimera. The results indicated that the homology modeling approach effectively generated reliable 3D models of GaHP2. The high confidence score (PROCHECK), stereochemical quality (VERIFY3D), energy of the protein chain (ANOLEA), and RMSD of 0.540 Å indicate that the proposed model closely reflects the actual protein conformation. One interesting finding of the study was the correlation between the presence of aromatic clusters in GaHP2 and its stability at higher temperatures, a finding not previously documented in cold-adapted Antarctic proteins. The GaHP2 protein was also found to contain domains that encoded oxygen-binding and/or oxygen-transporting globins, as indicated by functional analysis, suggesting a role in cold adaptation under low oxygen conditions. This research illustrates that thermal stress proteins may possess distinctive structural flexibility and stability that enable them to function under thermal stress, thereby safeguarding host organisms from heat aggregation and cold denaturation.*

INTRODUCTION

Glaciozyma antarctica is a native Antarctic yeast that has generated research interest due to its exceptional capacity to thrive in a diverse range of climatic extremes (Convey & Peck, 2019). Researchers have observed that *G. antarctica* can tolerate temperatures ranging from -12°C to 20°C (Boo *et al*., 2013; Soon *et al*., 2018). As a means of surviving and adapting to the Antarctic climate,

these extremophiles produce a wide range of biologically important proteins, particularly those involved in the thermal stress response (Song *et al*., 2017; Yusof *et al*., 2021). Genomic analysis of *G. antarctica* revealed several genes encoding proteins crucial to the capacity of the cell to adapt to cold settings (Firdaus-Raih *et al*., 2018). However, a considerable proportion of these proteins remain uncharacterized, particularly the conserved hypothetical proteins (HPs) (Soon *et al*., 2018).

Conserved HPs are proteins found in organisms from several phylogenetic lineages but have not been functionally characterized (Ijaq *et al*., 2019). Determining the structure and function of proteins shared by numerous organisms can provide insights into their evolutionary changes, operational mechanisms, and molecular functions (Hauri *et al*., 2019; Morris *et al*., 2022; Zhao *et al*., 2021). These functionally unknown proteins may be involved in significant aspects of this microorganism's biological function. For instance, the evaluation of the immune protective effects of EtCHP18905, a conserved hypothetical protein of the obligate intracellular parasite *Eimeria tenella*, revealed that it could be an effective candidate for the development of new vaccines (Zhao *et al*., 2021). Previous research has demonstrated that a set of proteins with unknown functions is vital in the physiological regulation and cold adaptation of psychrophilic microorganisms (Teoh *et al*., 2021). Another study suggests that *G. antarctica*'s HPs play a crucial role in the early stages of cold and freeze stress, although validation of these findings is still pending (Wong *et al*., 2019). This offers a chance to uncover new insights into the unique characteristics of their adaptation mechanisms, specifically regarding protein flexibility, structural stability, and residue interactions that facilitate functionality in cold and thermally fluctuating conditions.

The main problem in the structural determination of uncharacterized proteins is the difficulty of obtaining soluble proteins for downstream processing (Ahmad *et al*., 2018; Kielkopf *et al*., 2021). To understand biological processes at the system level, we must evaluate the three-dimensional (3D) protein structures that mediate biochemical interactions (Hauri *et al*., 2019). Structure determination experiments are also labor-intensive and complicated processes (Rigden, 2017). Comparative homology modeling, in the absence of an experimentally determined structure, can generate a useful 3D model of a protein that links to at least one known protein structure. The rapid improvements in omics technologies allow researchers to complement the expensive and time-consuming structural determination experiments. This has allowed the theoretical assessment of several physicochemical parameters to reveal the function and makeup of proteins about which we previously lacked information (Jumper *et al*., 2021).

Thus, by combining structural knowledge of proteins with functional annotation tools, previously uncharacterized proteins can be elucidated (Jez, 2017). Therefore, this study uses in silico analysis to investigate the structure–function relationship of heat shock proteins (HPs) involved in *G. antarctica*'s thermal stress response. The current study aims to evaluate the conserved HPs associated with thermal stress responses in *G. antarctica* to establish a better understanding of their adaptation mechanisms.

MATERIALS AND METHODS

Sequence Analysis

The *G. antarctica* PI12 genome transcriptome data (PRJNA41257) were examined for conserved thermal stress response heat shock proteins. Proteins suitable for protein structure determination were identified based on a reliable level of expression (1.5-fold, p-value < 10^{-5}) (Firdaus-Raih *et al*., 2018) and a protein size range of 18 to 55 kDa (Jez, 2017; Klebe, 2013). Normalized counts underwent differential expression analysis using DESeq2. Genes were deemed significantly differentially expressed if they exhibited an adjusted p-value of less than 0.05 and a log2 fold change of greater than or equal to ±1. The selected heat shock protein's amino acid FASTA sequence was translated from the DNA sequence through the ExPASy Translate Tool (https://web.expasy.org/translate/) and utilized for functional annotation, physicochemical analysis, and homology modeling. The sequence quality was

confirmed through sequence alignment and validation using the EMBL-EBI Clustal Omega Tool (Madeira *et al*., 2024).

Physicochemical Analysis

Physicochemical characteristics of HPs in raw sequence format were assessed using ExPASy's ProtParam tool (http://web.expasy.org/protparam/) (Gasteiger *et al*., 2005). The computed isoelectric point (pI) is valuable for protein characterization since it indicates that the surface of the protein is charged, but the overall charge of the protein is zero, rendering it stable and compact (Tokmakov *et al*., 2021).

Functional Annotation

An analysis was conducted for functional annotation, utilizing the Basic Local Alignment Search Tool (BLAST) to search for similar proteins in the NCBI non-redundant (nr) database. The purpose was to identify homologous proteins from related organisms that are likely to have the same function as the query protein. The BLAST search utilized an E-value threshold of 1e-5 and a maximum of 100 target sequences to ensure relevant and statistically significant matches. The presence of domains and important sites in functional protein families was predicted using InterProScan (https://www.ebi.ac.uk/interpro/about/interproscan), as described by Mitchell *et al*. (2019).

Comparative Homology Modelling and Structure Assessment

Structure prediction

Homology modeling was performed using both the Phyre2 server (www.sbg.bio.ic.ac.uk/phyre2) (Kelley *et al*., 2015) and the AlphaFold server (https://alphafoldserver.com/) (Abramson *et al*., 2024; Jumper *et al*., 2021). The 3D models generated were refined using the ModRefiner web server (https://zhanggroup.org/ModRefiner/). Refining the structure is an essential process in bringing a starting structure closer to its native state and achieving accuracy that is comparable to experimental results (Bhattacharya, 2019).

Structure validation

The best homology models were assessed for their stereochemical quality using the Ramachandran plot (Lovell *et al*., 2003) and VERIFY3D (https://servicesn.mbi.ucla.edu/Verify3D) (Eisenberg *et al*., 1997). The energy of the protein chain was calculated using the atomic empirical mean force potential ANOLEA (www.fundp.ac.be/pub/ANOLEA.html), as described by Melo *et al*. (1997). High-energy zones identified by ANOLEA denote regions of the protein exhibiting unfavorable atomic interactions, potentially corresponding to structurally strained, flexible, or poorly folded areas. These zones may indicate potential sites of instability or conformational variability, and in this study, they were interpreted as regions that could contribute to the dynamic nature or functional flexibility of the cold-adapted protein.

Once the 3D model was generated, the PDB files containing the constructed 3D models of *G. antarctica* HP were submitted to the DALI server for template search (Holm, 2020). Selected templates were incorporated with the 3D model of interest for comparative analysis using UCSF Chimera (Pettersen *et al*., 2004).

Protein-protein interactions were analyzed using the Protein Interactions Calculator (PIC) server (http://pic.mbu.iisc.ernet.in/), as described by Tina *et al*. (2007). Using the coordinate set of a protein or assembly, the PIC server calculates a range of interactions within the protein or complex. These include disulfide bonds, interactions between hydrophobic residues, ionic interactions, hydrogen bonds, aromatic–aromatic interactions, aromatic–sulfur interactions, and cation-π interactions.

RESULTS AND DISCUSSIONS

GaHP2 Sequence Analysis

The transcriptomic analysis of the *G. antarctica* genome revealed GaHP2, a 765 bp gene coding for a conserved hypothetical protein related to thermal stress response (Table 1). The gene exhibited a level of expression at low temperatures (below 37°C) greater than 1.5 (significant p-value 10^{-5}). The temperature threshold is significant as cold-adapted organisms, like *G. antarctica*, thrive in environments significantly below mesophilic conditions. Increased gene expression at temperatures below 37°C suggests a potential role in cold adaptation and survival mechanisms (Teoh *et al*., 2021). The verified sequences of the genes coding for GaHP2 were converted into their amino acid sequences via Expasy Translate Tools. Functional annotation and physicochemical analysis of the protein were conducted using the amino acid sequences arranged in GenBank/GB format. Functional annotation employs peptide sequences, considering that protein amino acid sequences are more conserved than gene nucleotide sequences (Kapli *et al*., 2023).

Table 1. Sequence analysis of *G. antarctica* transcriptomic data for conserved hypothetical proteins related to thermal stress response.

Transcript	Length (bp)	Protein ID	Description
GaHP2	765	GaHP2	Predicted uncharacterized protein, conserved

Physicochemical Analysis

ExPASy's ProtParam tool revealed the molecular weight of GaHP2 at 29 kDa and isoelectric point (pI) values of 5.46, indicating that the protein is acidic and negatively charged. Proteins exhibit a wide range of physicochemical properties that significantly impact their activity, structure, and consequently, their biological function (Yuan *et al*., 2020). Understanding the function and composition of HPs can be enhanced by computing and predicting their physicochemical properties (Moldoveanu & David, 2022; Naqvi *et al*., 2015). Understanding the pI value is crucial in protein purification as it reveals the pH value at which solubility tends to be at its lowest.

The protein pI represents the specific pH value at which the protein's movement comes to a halt in an electro-focusing device. This pH value also indicates the point at which the protein will be eluted (Novák & Havlíček, 2016). In addition, the molecular weight and pI parameters can be used to analyze two-dimensional electrophoresis gels, which helps in the experimental analysis of proteins (Molina-Mora *et al*., 2020; Sahab *et al*., 2005). Biologists have previously utilized the theoretical and experimental determination of the isoelectric point of peptides to assist in identifying peptides in complex mixtures (Kozlowski, 2021). With the help of the computed pI, protein purification protocols and crystal screening strategies can be developed.

Functional Annotation

After conducting a Protein BLAST search against the NCBI non-redundant (nr) database, it was found that GaHP2 has a similarity of less than 57% to the protein in the database (Table 2). Sequences were considered homologous only if they exhibited a 90% identity in a BLAST search against the NCBI non-redundant (nr) database (Altschul *et al*., 1997; Gazi *et al*., 2020). Based on the findings, it appears that the protein's characterization is still incomplete.

Table 2. The BLAST search results for GaHP2, a conserved hypothetical protein from *G. antarctica*, against the NCBI non-redundant (nr) database.

Protein ID	Sequence identity	e value	Description
GaHP2	57%	7.00E-86	ORY79343.1 Protoglobin-domain-containing protein [*Leucosporidium creatinivorum*]

InterProScan also provided limited information on known proteins for predicting homologous protein families and functional domains (Table 3). The GaHP2 protein appeared to be associated with biological processes related to oxygen and heme binding. However, no specific functional domain could be predicted for this protein. It is evident that there is still much to learn about GaHP2, allowing for potential novel discoveries. Described as part of the Globin or protoglobin homologous superfamily, GaHP2 also exhibits two GO terms for biological processes: heme binding and oxygen binding. It appears that this protein plays a role in the transportation of oxygen, similar to how haemoglobin functions in humans. Previously, scientists found that *G. antarctica* PI12 displayed remarkable adaptability in low oxygen conditions. By utilizing nitrite as an alternative electron terminal acceptor, along with other common mechanisms, *G. antarctica* PI12 was able to thrive in the cold environment (Wong *et al*., 2019).

Table 3. Functional annotation of GaHP2, a conserved hypothetical protein from *G. antarctica*, using InterProScan tool.

Protein ID	Homologous Superfamily	Domains	Biological process	Molecular function	Cellular component
GaHP2	Globin/Proto (IPR012292)	None predicted	heme binding (GO:0020037) oxygen binding (GO:0019825)	None predicted	None predicted

3D Model Development

The 3D structures of the GaHP2 protein exhibit a prominent pattern of α-helices, along with a significant number of random loops in the region between the α-helices and β-sheet (Figure 1). Based on the abundance of looped regions, it can be inferred that the protein structure is remarkably conserved (Neelamathi *et al*., 2009). The presence of numerous looped regions suggests that, despite a low overall sequence identity (< 30%), the tertiary structure topology, especially the fold and loop-rich architecture, might be conserved. This indicates that structural motifs or functional topologies, rather than primary sequences, may be retained to ensure stability or function in cold-adapted environments (Hamid *et al*., 2022; Michetti *et al*., 2017).

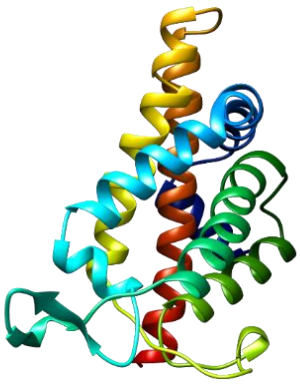


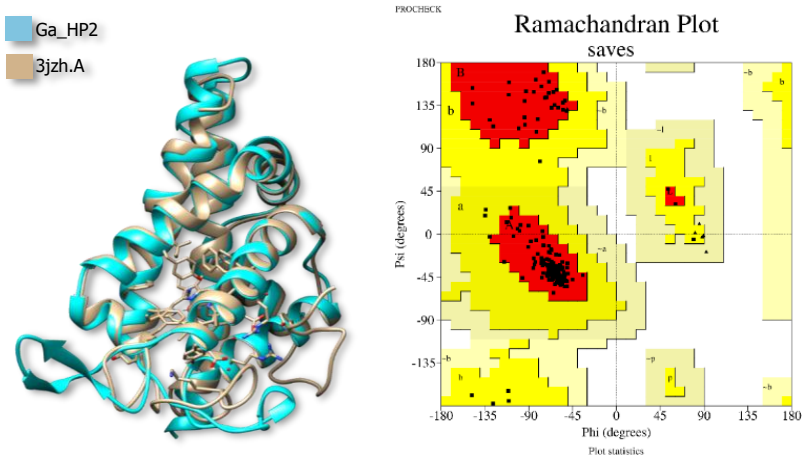
Figure 1. The three-dimensional (3D) structure of the GaHP2 protein constructed using Phyre2 Server.

The templates used by Phyre2 server for the 3D structure construction were Protoglobin (c2veeC_) from *Methanosarcina acetivorans* with 15% sequence identity and 100% confidence score. Confidence score represents the probability (from 0 to 100) that the match between our sequence and its template is a true homology. The score of > 90% indicates protein adopts the overall fold shown and that the core of the protein is modelled at high accuracy (Kelley *et al.*, 2015). The confidence score indicated that GaHP2 models fulfilled the structural validation score requirement despite a sequence identity of less than 30% similar to PDB structures. A high degree of confidence (> 90%) and a low degree of sequence identity (< 30%) suggest that the fold is likely right, accurate in the core (2 - 4Å), but may exhibit significant variations in loops and noncore areas. Phyre2 combines PSI-BLAST and the Hidden Markov Model (HMM) database to provide an extremely strong protein structure prediction algorithm capable of reliably identifying very distant homology and creating accurate models even when sequence identity is less than 15% (Kelley *et al.*, 2015). Additionally, structure refinement through ModRefiner can minimise template structure errors related to sequence identity below 25% (Adiyaman & McGuffin, 2019).

Comparative structural modelling is reliable and can predict protein structures with atomic precision (Jumper *et al.*, 2021). The 3D structure was also compared with the latest AlphaFold2 (Abramson *et al.*, 2024; Jumper *et al.*, 2021) and RoseTTAfold (Baek *et al.*, 2021) programs and results in less than 1.4 Å. The computational method for protein structure determination has advanced to the point that it can predict protein structures with atomic precision regularly, even in cases where no similar structure exists (Jumper *et al.*, 2021). This provides a reliable and cost-effective alternative to the months to years of laborious effort required to determine the structure of a single protein.

Model Validation and Assessment

The structural comparisons of the constructed 3D model with its homologs, as well as the corresponding Ramachandran plot statistics, were presented in Figure 2. The PROCHECK through Ramachandran plot analysis revealed that GaHP2 model had a good quality score, with no residues in disallowed regions being found. DALI server was able to retrieve compatible homologs for GaHP2 for comparative structure analysis. Using the UCSF Chimera programme, structural comparisons of GaHP2 3D models resulted in a 14% overlap. The summary of structural quality evaluation shown in Table 4 confirmed that GaHP2 model was acceptable, with an RMSD value of 0.540 Å indicating very little variation from homologs.



(a) (b)

Figure 2. GaHP2 structural comparison with templates obtained from the DALI server: (a) Superimposition of GaHP2 predicted structure (turquoise) and hEED (PDB ID: 3jzh) (tan) with RMSD value 0.5 Å; and (b) Ramachandran plot provided by the PROCHECK program for the superimposed GaHP2 model with a comparison with 3jzh.

Table 4. Summary of structural quality assessment results of the superimposed 3D Structure of GaHP2 with templates retrieved from DALI server.

3D Model	% Sequence identity	PROCHECK V.3.5 (expected > 90% in favored region)	Verify3D (expected > 80%)	ANOLEA (expected < 35%)	RMSD
GaHP2	14% 3jzh.B Protoglobin *Methanosarcina acetivorans*	100.0% allowed 0.0% disallowed	91.62%	36.84% Z-score: 4.43	0.540 Å

Additionally, the 3D model achieved an excellent Verify3D score of 91.62%. Verify3D evaluates protein structures using 3D profiles. This software determines if an atomic model (3D) is compatible with its amino acid sequence (1D) (Eisenberg *et al*., 1997). The 3D-1D scoring matrix is a 20 × 18 matrix that reflects the environmental compatibility of each residue. The higher the value, the better the entire model's compatibility (Matsuo *et al*., 1995). For ANOLEA's energy calculations, it is noted that the GaHP2 built 3D model scored 36.84% in the high energy zone (HEz) in the protein profile, slightly higher than the acceptable score of 35%. However, a closer observation revealed that a significant portion of the amino acids with a high energy content were in the non-superimposed loop (Figure 3). A previous study showed that ANOLEA energy assessment of a few loop regions was high with positive values even after undergoing loop refinement (Pan *et al*., 2021; Singh *et al*., 2012). This is corroborated by ANOLEA's low Z-score of 4.43 for the GaHP2 model. The Z-score is calculated using the pseudo energies of target protein sequences, with a lower Z-score indicating more reliability (Melo & Feytmans, 1998). This is verified by the fact that the HEz value of 36.95% obtained from the 3D model of Arabidopsis thaliana HAC1 protein profile correlates with structural defects, while the lower Z-score of 3.51 suggests a high-quality 3D model (Ćemanović *et al*., 2014).

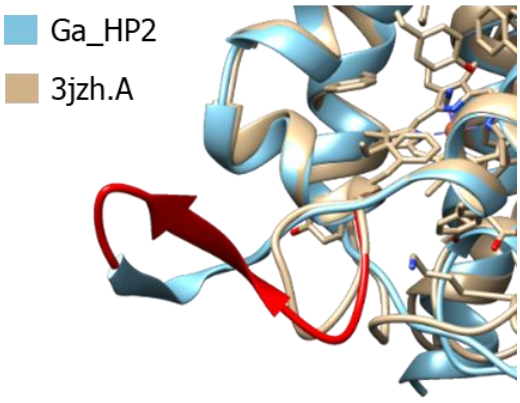


Figure 3. The amino acids in GaHP2's non-superimposed loop with high energy content are highlighted in red.

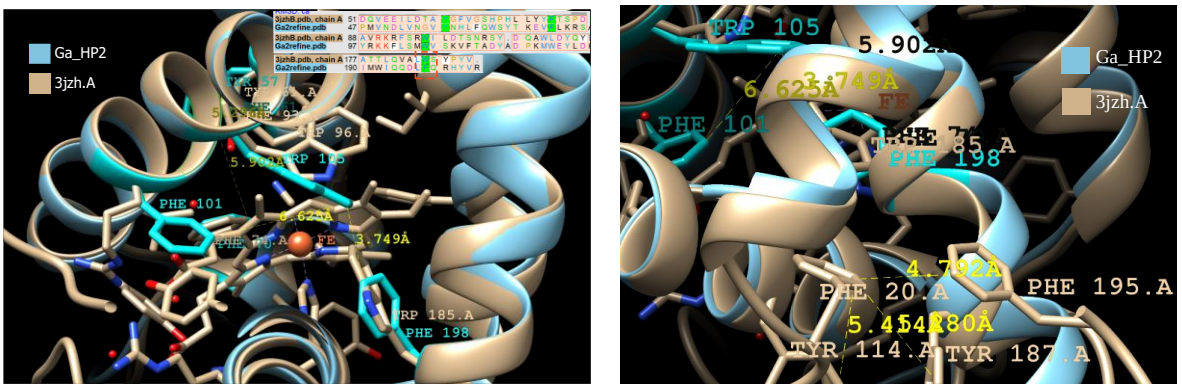
Analysis of the intra-protein interaction of the constructed 3D *G. antarctica* model with its respective homologs is shown in Table 5. Generally, GaHP2 3D models possessed lower hydrophobic interactions compared to their homologs. This demonstrates their structural flexibility, which enables them to function at low temperatures (Hamid *et al*., 2022). Increased structural flexibility of specific residues, mainly in the active region, or the entire protein structure, enables adaptation of psychrophilic enzymes by lowering the activation energy required for metabolism (Fields, 2001). The findings are corroborated by a study of a cold-adapted protein, chitobiase from *Arthrobacter sp.* TAD20, in which increased structural flexibility resulted in increased activity at low temperatures, particularly those that significantly impede molecular motions (Hamid *et al.*, 2022). A previous study on cold-adapted β-d-Galactosidases from *Arthrobacter sp.* 32cB demonstrated that the environment's lack of free energy, caused by low temperature and high

viscosity of water, can be compensated for by a higher efficiency of energy gain due to their high structural flexibility (Rutkiewicz *et al.*, 2019).

Table 5. Comparative analysis of intra- and inter-protein interaction between GaHP2 and its homologs, protoglobin (PDB ID: 3zjh chain B) from the mesophilic *Methanosarcina acetivorans*, the globin domain of a globin-coupled sensor protein (PDB ID: 2w31 chain A) from *Geobacter sulfurreducens*, and the heme-based aerotactic transducer (HemAT) sensor domain (PDB ID: 1or4 chain A) from *Bacillus subtilis*.

Protein Interaction	GaHP2	3zjh-B	2w31-A	1or4-A
Hydrophobic Interaction within 5 Å	135	166	145	145
Ionic Interactions within 6 Å	18	18	19	19
Aromatic-Aromatic Interactions within 4.5 and 7 Angstroms	7	20	2	2

In contrast, the higher aromatic-aromatic interaction in GaHP2 as compared to the two mesophilic homologs, the globin domain (2w31) from *G. sulfurreducens* and the HemAT sensor domain (1or4) from *B. subtilis*, indicated unusual rigidity for a psychrophilic protein. However, further inspection showed that the aromatic-aromatic interaction in GaHP2 was located at the heme-binding site (Figure 4a). This indicates that the aromatic-aromatic interaction may contribute to the high binding site efficiency. Some protein regions that are not involved in catalysis may become even more rigid than their mesophilic counterparts, as previously described in the comparison of psychrophilic and mesophilic DNA ligases' 3D models (Georlette *et al.*, 2003). This correlates with the presence of a significant aromatic-aromatic interaction distant to the heme-binding site of the protoglobin (3zjh) homolog from the obligate anaerobic Archaea, *Methanosarcina acetivorans*, therefore contributing to the rigidity of the protein (Klinger *et al.*, 2003; Mitra & Das Mohapatra, 2021). In contrast, the protoglobin (3zjh) homolog was also detected with additional aromatic-aromatic interaction away from the heme-binding site (Figure 4b), therefore contributing to their overall structural rigidity. This was correlated with a previous report indicating that the obligate anaerobic Archaea, *M. acetivorans*, may be thermophilic facultatively (Klinger *et al.*, 2003).



(a) (b)

Figure 4. The aromatic-aromatic interaction takes place at heme binding sites in GaHP2 and its protoglobin (3zjh) homolog: (a) The aromatic residues are generally conserved except for some substitution of TRP185 in 3zjh homolog to PHE198 in GaHP2; and (b) Strong aromatic-aromatic interaction between PHE20 to TYR114, PHE195, and TYR187 in the protoglobin (3zjh) homolog from *Methanosarcina acetivorans*.

This finding, in conjunction with a previous study comparing the cold-adapted citrate synthase from an *Arthrobacter* strain to the homologous enzyme from the hyperthermophile *Pyrococcus furiosus*, indicates that part of the cold-adapted enzyme's adaptation to low temperature may be due to increased accessibility of the active site (Gerike *et al.*, 2001; Liu *et al.*, 2023). Due to the high degree of

conservation and relative rigidity of the catalytic site across homologous enzymes, the GaHP2 binding efficiency is consistent with the capacity of the *M. acetivorans* protoglobin (3zjh) homolog to reshape its haem distal site structure. By changing the conformation of another aromatic residue, Trp60, the aromatic residues Phe93 operate as ligand sensors and control access to the haem through the tunnel system (Klinger *et al*., 2003). In addition, the decreased solubility of oxygen as temperatures rise requires the active sites to be more efficient at a higher temperature than their optimal growth temperature (Rafiq *et al*., 2019). Finally, the reversibility of some cold-adapted enzymes' thermal unfolding has been attributed to the fact that when they are unfolded, a smaller number of hydrophobic groups are exposed to the aqueous solvent, thereby preventing or limiting the irreversible aggregation process typical of more stable proteins (Nowak & Otzen, 2024). The structural insights derived from this study may be utilized to design enzymes with specific characteristics for biotechnological applications. Furthermore, expanding comparative studies to include a wider range of psychrophilic, mesophilic, and thermophilic organisms may uncover additional evolutionary adaptations that optimize thermostability or flexibility in proteins similar to GaHP2.

CONCLUSION

This work presents the identification and in silico analysis of GaHP2, a conserved hypothetical protein that was involved in the thermal stress response in *G. antarctica*. The evaluation of physicochemical properties aided in understanding the unique characteristics of the annotated proteins, while functional annotation highlighted the role of the proteins in the biological processes of heme binding and oxygen binding. Comparative structural analysis of GaHP2 indicated cold-adapted traits, most notably increased flexibility in comparison to their mesophilic or thermophilic counterparts. This mostly results from their reduced hydrophobic, ionic, and aromatic interactions. Furthermore, the HPs include multiple additional loops and glycine residues, which have been associated with enhanced flexibility for activities suited to low temperatures. However, the presence of aromatic clusters in GaHP2 has been linked to the psychrophilic protein's unusually high thermostability. Therefore, this protein was hypothesized to preserve an ideal balance between molecular stability and structural adaptability in order to optimize its functional roles in oxygen transportation under thermal stress conditions. Further research should focus on experimental validation via in vitro enzyme assays, site-directed mutagenesis, and protein crystallography to verify the specific function of GaHP2 in the adaptation mechanism of *G. antarctica*.

ACKNOWLEDGEMENT

The research received funding from the Yayasan Penyelidikan Antartika Sultan Mizan (YPASM) grant (Ref UMS Code: GLS0032-2019), which serves as the main source of financial support for the project. In addition, authors would like to express gratitude for the support received from the Malaysian Ministry of Higher Education (MOHE), grant number FRGS/1/2017/STG05/UMS/02/2 and registered in UMS with the code number FRGS0463-2017.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E. & Jumper, J. M. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Adiyaman, R. & McGuffin, L.J. 2019. Methods for the refinement of protein structure 3D models. *International Journal of Molecular Sciences*, 20(9), 2301. https://doi.org/10.3390/ijms20092301.

Ahmad, I., Nawaz, N., Darwesh, N.M., ur Rahman, S., Mustafa, M.Z., Khan, S.B. & Patching, S.G. 2018. Overcoming challenges for amplified expression of recombinant proteins using *Escherichia coli*. *Protein Expression and Purification*, 144(November 2017), 12–18. https://doi.org/10.1016/j.pep.2017.11.005.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., Van Dijk, A.A., Ebrecht, A.C. & Baker, D. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871–876. https://doi.org/10.1126/science.abj8754.

Bhattacharya, D. 2019. RefineD: Improved protein structure refinement using machine learning based restrained relaxation. *Bioinformatics*, 35(18), 3320–3328. https://doi.org/10.1093/bioinformatics/btz101.

Boo, S. Y., Wong, C.M.V.L., Rodrigues, K.F., Najimudin, N., Murad, A.M.A., Mahadi, N.M., Yee, S., Clemente, B., Vui, M., Kenneth, Rodrigues, F., Munir, A. & Mahadi, M. 2013. Thermal stress responses in Antarctic yeast, *Glaciozyma antarctica* PI12, characterized by real-time quantitative PCR. *Polar Biology*, 36(3), 381–389. https://doi.org/10.1007/s00300-012-1268-2.

Ćemanović, A., Šutković, J., Gawwad, M.R.A., Cemanovic, A., Sutkovic, J. & Gawwad, M.R.A. 2014. Comparative structural analysis of HAC1 in *Arabidopsis thaliana*. *Network Biology*, *4*(2), 67–73. https://doi.org/10.0000/issn-2220-8879-networkbiology-2014-v4-0006.

Convey, P. & Peck, L.S. 2019. Antarctic environmental change and biological responses. *Science Advances*, 5(11). https://doi.org/10.1126/sciadv.aaz0888.

Eisenberg, D., Lüthy, R. & Bowie, J.U. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, 277, 396–404.

Fields, P.A. 2001. Protein function at thermal extremes: Balancing stability and flexibility. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 129(2–3), 417–431.

Firdaus-Raih, M., Hashim, N.H.F., Bharudin, I., Bakar, M.F.A., Huang, K.K., Alias, H., Lee, B.K.B., Isa, M.N.M., Mat-Sharani, S., Sulaiman, S., Tay, L.J., Zolkefli, R., Noor, Y.M., Law, D.S.N., Rahman, S.H.A., Md-Illias, R., Bakar, F.D.A., Najimudin, N., Murad, A.M.A. & Mahadi, N.M. 2018. The *Glaciozyma antarctica* genome reveals an array of systems that provide sustained responses towards temperature variations in a persistently cold habitat. *PLoS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0189947.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. & Bairoch, A. 2005. *The Proteomics Protocols Handbook*. 571–608. https://doi.org/10.1385/1592598900.

Gazi, M.A., Mahmud, S., Fahim, S.M., Md. Rezaul, I., Das, S., Mahfuz, M. & Ahmed, T. 2020. Questing functions and structures of hypothetical proteins from *Campylobacter jejuni*: A computer-aided approach. *Bioscience Reports*, 40(6), 1–14. https://doi.org/10.1042/BSR20193939.

Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V.N., Gerday, C. & Feller, G. 2003. Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. *Journal of Biological Chemistry*, 278(39), 37015–37023. https://doi.org/10.1074/jbc.M305142200.

Gerike, U., Danson, M.J. & Hough, D.W. 2001. Cold-active citrate synthase: Mutagenesis of active-site residues. *Protein Engineering*, 14(9), 655–661.

Hamid, B., Bashir, Z., Yatoo, A.M., Mohiddin, F., Majeed, N., Bansal, M., Poczai, P., Almalki, W.H., Sayyed, R.Z., Shati, A.A. & Alfaifi, M.Y. 2022. Cold-active enzymes and their potential industrial applications - a review. *Molecules*, 27(18), 5885. https://doi.org/10.3390/molecules27185885.

Hauri, S., Khakzad, H., Happonen, L., Teleman, J., Malmström, J. & Malmström, L. 2019. Rapid determination of quaternary protein structures in complex biological samples. *Nature Communications*, 10(1), 1–10. https://doi.org/10.1038/s41467-018-07986-1.

Holm, L. 2020. DALI and the persistence of protein shape. *Protein Science*, 29(1), 128–140. https://doi.org/10.1002/pro.3749.

Ijaq, J., Malik, G., Kumar, A., Das, P.S., Meena, N., Bethi, N., Sundararajan, V.S. & Suravajhala, P. 2019. A model to predict the function of hypothetical proteins through a nine-point classification scoring schema. *BMC Bioinformatics*, 20(1), 1–8. https://doi.org/10.1186/s12859-018-2554-y.

Jez, J.M. 2017. Revisiting protein structure, function, and evolution in the genomic era. *Journal of Invertebrate Pathology*, 142, 11–15.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J. & Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kapli, P., Kotari, I., Telford, M.J., Goldman, N. & Yang, Z. 2023. DNA sequences are as useful as protein sequences for inferring deep phylogenies. *Systematic Biology*, 72(5), 1119–1135. https://doi.org/10.1093/sysbio/syad036.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6), 845–858. https://doi.org/10.1038/nprot.2015.053.

Kielkopf, C.L., Bauer, W. & Urbatsch, I.L. 2021. Expression of cloned genes in *E. coli* using IPTG-inducible promoters. *Cold Spring Harbor Protocols*, 2021(2), pdb-prot102137.

Klebe, G. 2013. Experimental methods of structure determination. In G. Klebe (Ed.), *Drug Design* (pp. 265–290). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17907-5_13.

Klinger, C., Robbach, M., Howe, R. & Kaufmann, M. 2003. Thermophile-specific proteins: The gene product of aq_1292 from *Aquifex aeolicus* is an NTPase. *BMC Biochemistry*, 4, 1–12. https://doi.org/10.1186/1471-2091-4-12.

Kozlowski, L.P. 2021. IPC 2.0: Prediction of isoelectric point and p*K* dissociation constants. *Nucleic Acids Research*, 49(W1), W285–W292. https://doi.org/10.1093/nar/gkab295.

Liu, Y., Zhang, N., Ma, J., Zhou, Y., Wei, Q., Tian, C. & Zhang, S. 2023. Advances in cold-adapted enzymes derived from microorganisms. *Frontiers in microbiology*, 14, 1152847.

Lovell, S.C., Davis, I.W., Arendall III, W.B., De Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S. & Richardson, D.C. 2003. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3), 437–450.

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. & Butcher, S. 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic acids research, 52(W1), W521-W525.

Matsuo, Y., Nakamura, H. & Nishikawa, K. 1995. Detection of protein 3D-1D compatibility characterized by the evaluation of side-chain packing and electrostatic interactions. *The Journal of Biochemistry*, 118(1), 137–148.

Melo, F., Devos, D., Depiereux, E. & Feytmans, E. 1997. ANOLEA: a www server to assess protein structures. *International Conference on Intelligent Systems for Molecular Biology*, 5, 187–190.

Melo, F. & Feytmans, E. 1998. Assessing protein structures with a non-local atomic interaction energy. *Journal of Molecular Biology*, 277(5), 1141–1152. https://doi.org/10.1006/jmbi.1998.1665.

Michetti, D., Brandsdal, B.O., Bon, D., Isaksen, G.V., Tiberti, M. & Papaleo, E. 2017. A comparative study of cold and warm-adapted Endonucleases A using sequence analyses and molecular dynamics simulations. *PloS One*, 12(2), e0169586.

Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H. & Finn, R.D. 2019. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351–D360. https://doi.org/10.1093/nar/gky1100.

Mitra, D. & Das Mohapatra, P. K. 2021. Cold adaptation strategy of psychrophilic bacteria: an in-silico analysis of isocitrate dehydrogenase. *Systems Microbiology and Biomanufacturing*, 1(4), 483-493.

Moldoveanu, S. & David, V. 2022. Characterization of analytes and matrices. In *Essentials in Modern HPLC Separations* (pp. 179–205). Elsevier. https://doi.org/10.1016/B978-0-323-91177-1.00003-X

Molina-Mora, J.A., Chinchilla-Montero, D., Castro-Peña, C. & García, F. 2020. Two-dimensional gel electrophoresis (2D-GE) image analysis based on CellProfiler: *Pseudomonas aeruginosa* AG1 as model. *Medicine*, *99*(49), e23373. https://doi.org/10.1097/MD.0000000000023373.

Morris, R., Black, K.A. & Stollar, E.J. 2022. Uncovering protein function: From classification to complexes. *Essays in Biochemistry*, 66(3), 255–285. https://doi.org/10.1042/EBC20200108.

Naqvi, A.A.T., Shahbaaz, M., Ahmad, F. & Hassan, M.I. 2015. Identification of functional candidates amongst hypothetical proteins of *Treponema pallidum* ssp. *Pallidum. PloS One*, *10*(4), e0124177. https://doi.org/10.1371/journal.pone.0124177.

Neelamathi, E., Vasumathi, E., Bagyalakshmi, S. & Kannan, R. 2009. Insilico prediction of structure and functional aspects of a hypothetical protein of *Neurospora crassa*. *Journal of Cell and Tissue Research*, 9(3), 1989.

Novák, P. & Havlíček, V. 2016. *Protein extraction and precipitation*. *proteomic profiling and analytical chemistry: the crossroads*. Second Edition, 52–62. https://doi.org/10.1016/B978-0-444-63688-1.00004-5.

Nowak, J.S. & Otzen, D.E. 2024. Helping proteins come in from the cold: 5 burning questions about cold-active enzymes. *BBA Advances*, *5*, 100104. https://doi.org/10.1016/j.bbadva.2023.100104.

Pan, F., Zhang, Y., Lo, C.C., Mandal, A., Liu, X. & Zhang, J. 2021. *Protein loop modeling and refinement using deep learning models*. https://doi.org/10.1101/2021.11.03.467148.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. 2004. UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. https://doi.org/10.1002/jcc.20084.

Rafiq, M., Hassan, N., Rehman, M. & Hasan, F. 2019. Adaptation mechanisms and applications of psychrophilic fungi. In *Fungi in Extreme Environments: Ecological Role and Biotechnological Significance* (pp. 157–174). Springer.

Rigden, D.J. 2017. From protein structure to function with bioinformatics: Second Edition. In *From Protein Structure to Function with Bioinformatics*. Second Edition. https://doi.org/10.1007/978-94-024-1069-3.

Rutkiewicz, M., Bujacz, A. & Bujacz, G. 2019. Structural features of cold-adapted dimeric GH2 β-D-galactosidase from *Arthrobacter* sp. 32cB. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1867*(9), 776–786. https://doi.org/10.1016/j.bbapap.2019.06.001.

Sahab, Z.J., Suh, Y. & Sang, Q.X.A. 2005. Isoelectric point-based prefractionation of proteins from crude biological samples prior to two-dimensional gel electrophoresis. *Journal of Proteome Research*, 4(6), 2266–2272.

Singh, S.P., Vivek, S., Bezbaruah, R. L. & Barooah, M. 2012. Prediction of the three-dimensional structure of serine/threonine protein kinase pto of *Solanum lycopersicum* by homology modelling. *Bioinformation*, *8*(5), 212–215. https://doi.org/10.6026/97320630008212.

Song, Y.S., Seo, D.J. & Jung, W.J. 2017. Identification, purification, and expression patterns of chitinase from psychrotolerant *Pedobacter* sp. PR-M6 and antifungal activity in vitro. *Microbial Pathogenesis*, 107, 62–68. https://doi.org/10.1016/J.MICPATH.2017.03.018.

Soon, J., Koh, P., Vui, C. M., Wong, C. M.V.L., Najimudin, N. & Mahadi, M. 2018. Gene expression patterns of *Glaciozyma antarctica* PI12 in response to cold, and freeze stress. *Polar Science*, 20(1): 45-54. https://doi.org/10.1016/j.polar.2018.11.007.

Teoh, C.P., Lavin, P., Lee, D.J.H., González-Aravena, M., Najimudin, N., Lee, P.C., Cheah, Y.K., & Wong, C.M.V.L. 2021. Genomics and transcriptomics analyses provide insights into the cold adaptation strategies of an Antarctic bacterium, *Cryobacterium sp*. SO1. *Polar Biology*, *44*(7), 1305–1319. https://doi.org/10.1007/s00300-021-02883-8.

Tina, K.G., Bhadra, R. & Srinivasan, N. 2007. PIC: Protein Interactions Calculator. *Nucleic Acids Research*, 35(SUPPL.2), 473–476. https://doi.org/10.1093/nar/gkm423.

Tokmakov, A.A., Kurotani, A. & Sato, K.I. 2021. Protein pI and intracellular localization. *Frontiers in Molecular Biosciences*, *8*, 775736. https://doi.org/10.3389/fmolb.2021.775736.

Wong, C.M.V.L., Boo, S.Y., Voo, C.LY., Zainuddin, N. & Najimudin, N. 2019. A comparative transcriptomic analysis provides insights into the cold-adaptation mechanisms of a psychrophilic yeast, *Glaciozyma antarctica* PI12. *Polar Biology*, 42(3), 541–553. https://doi.org/10.1007/s00300-018-02443-7.

Yuan, H., Lan, P., He, Y., Li, C. & Ma, X. 2020. Effect of the modifications on the physicochemical and biological properties of β-glucan—A critical review. *Molecules*, 25(1), 57.

Yusof, N.A., Hashim, N.H.F. & Bharudin, I. 2021. Cold adaptation strategies and the potential of psychrophilic enzymes from the Antarctic yeast, *Glaciozyma antarctica* pi12. *Journal of Fungi*, 7(7). https://doi.org/10.3390/jof7070528.

Zhao, H., Zhu, S., Zhao, Q., Huang, B., Liu, G., Li, Z., Wang, L., Dong, H. & Han, H. 2021. Molecular characterization and protective efficacy of a new conserved hypothetical protein of *Eimeria tenella*. *Parasite*, *28*. https://doi.org/10.1051/parasite/2021037.